

## ▶ ネットワーク情報化社会におけるアーカイブ ーデジタル情報環境における持続可能性ー

杉本 重雄

筑波大学図書館情報メディア研究科

### 概要

国、地方自治体ともにネットワークを基盤とする電子政府・電子自治体サービスやビジネスプロセスの開発のために相当な努力を重ねてきている。しかしながら、公文書や政府資料に関する情報収集プロセスを見るとデータの長期保存に関する懸念を持たざるを得ない。本稿では、2007年10月開催の第8回 EASTICA 総会で取り上げた話題を基礎にしてアーカイブの持続可能性に関する課題を考察する。今回の発表で扱った問題は、デジタルアーカイビングの基本モデル、デジタル資源のライフサイクル、そして、長期的なデジタル資源の利用のための保存ならびにメタデータ技術である。

### 1. はじめに

紙は、千年以上にも渡り情報を記録し文書を作成する主な媒体として利用されてきた。他方で、情報技術の発展により、私たちの暮らしの中の多くの場面で、紙はデジタルメディアに置き換えられてきた。現在、我々が目にする文書はほとんどすべてが電子的に作られ、また、その多くがネットワーク情報環境で配信されている。例えば、電子メールやウェブ資料を読むためだけに紙に印刷し、読んだ後は捨ててしまうことをしばしば目にする。これはある意味での「ペーパーレス」環境である。デジタル文書の利用により紙の消費量が減ったわけではないが、紙への依存度は小さくなっている。ここで懸念されるのは、将来電子文書を読むことができなくなるかもしれないということである。

デジタル文書の保存は良く知られた問題である。将来に備えて今ある情報を保存することが重要であることは分かってはいる。しかしながら、どうすれば本当に保存できるのかは十分には分かっていない。その主な理由は、保存が求められる期間と特定の資料の利用に必要なハードウェアとソフトウェアの寿命の違いである。さらには、デジタル文書や記録を保存用に適切かつ効果的に収集す

ることの難しさがある。

政府は業務上、実に多くの文書を作成し、大量の情報を発信している。このように政府が作り出す大量の文書や記録は、法律に従って保存していかなければならない。

電子政府環境では、我々を取り巻く社会の情報基盤に適合するように情報資源を作成し、発信し、利用できるようにしていかなければならない。そこでは、最先端の情報技術を採用していくことが求められることも多い。紙の文書を基礎とする従来の文書利用と管理の環境に比較して、一般に、高度なデジタル文書技術を用いることで柔軟かつ拡張可能な文書の利用と管理の環境を構築することができる。しかし、柔軟性や拡張可能性が、結局は、デジタル文書の保存を難しくしている

この数年、筆者は、国立国会図書館納本制度審議会のネットワーク出版物の収集に関する委員会、内閣府におけるデジタル媒体を利用した公文書の管理、移管、保存について研究する委員会など、各種委員会やグループに参加し、デジタル資源の保存について考える機会を多く得た。また、デジタル図書館やメタデータに関する研究の中でデジタル資源の長期保存について研究する機会も得た。本稿は、こうした経験と、2007年10月に東京で開

催された ICA 東アジア支部 (EASTICA) の第 8 回総会における討論をふまえたものである<sup>1</sup>。

## 2. デジタルアーカイブとデジタル文書

本稿では、「デジタルアーカイブ (digital archive)」を、収集されたデジタル資源を管理するための総合的なサービスと定義する。デジタルアーカイブには、文化遺産アーカイブ、学術論文のリポジトリ、政府や企業の記録アーカイブなど、様々なものがある。このように、デジタルアーカイブということばは広い意味で用いられるが、本稿では、政府資料を中心とするデジタル文書を収集保存するものや、文化遺産や学術資料を収集保存するものについて考察する。

デジタルアーカイブに保存されるデジタル資源は、ボーンデジタル資源 (born-digital resource) と非デジタル資源からデジタル化された資源 (Digitized Resource) の 2 種類に大別することができる。一方、デジタル化資源とボーンデジタル資源の境目は曖昧であり、本稿ではデジタル資源とボーンデジタル資源は区別しない。

本稿では、「文書 (document)」という用語を、情報を記録するために作成されるオブジェクトを表すために用いる。文書は紙資料かデジタル資料である。デジタル文書を実現するために使われるフォーマットには、プレーンテキスト、ワープロ文書、スプレッドシート、ハイパーテキスト、画像や音声が埋め込まれたマルチメディア文書など、様々なものがある。デジタル文書は CD や DVD などのパッケージ媒体あるいはインターネットを用いて発信・交換される。

デジタル文書は紙の場合とはかなり異なる動的性質を持っている。まずはじめに、デジタル文書は適当なハードウェアやソフトウェア無しに閲覧したり編集したりすることはできない。デジタル文書には、ハイパーリンク、テキスト検索、埋め

込み画像などの機能があり、これは紙の文書はもてない特性である。また、ワープロ文書から PDF へというように、目的や使用環境に応じてフォーマットの変換が頻繁に行われる。もう一つの重要な特徴は、ダイレクトメールの文書のようにデータを加工することで作り出されるデジタル文書があるという点である。こうした柔軟な特徴は文書の利用性を高めるには非常に役に立つ。一方、保存の場合には、保存に適した文書内容と形式を選ばなくてはならないという基本的な問題があることを意味する。

デジタルアーカイブには、オリジナルの資料を長期に渡って、元のまま保存し、利用に供することが期待される。しかし、デジタルアーカイブのサービスのどれもが長期的利用のために設計されているわけではない。例えば、ユーザインタフェースの変化や文書のマイグレーションのために、文書の見え方や使い勝手が変更される場合もある。

どんなコンテンツ (内容) を保存しなければならないかは、デジタルアーカイブの運営者が慎重に決めるべき問題である。デジタル保存のためのメタデータスキームとして開発された PREMIS<sup>2</sup> は、図 1 に示す興味深いデータモデルに基づいている。このモデルには知的内容、オブジェクト、権利、イベント、そして、エージェントの 5 つの実体がある。例えば、マイクロソフトの Word 文書ファイルや、その Word 文書ファイルから作られた PDF ファイルはいずれもオブジェクトであり、同一の知的内容に関連づけられている。デジタルアーカイブはデジタルオブジェクトのみならず知的内容の保存も求められる。文書の知的内容を保存することがアーカイブの第一の目標であり、ビットレベルの保存は二次的な目的であるという場合、知的内容を変えてしまうことにならない限り、オブジェクトを別の形式のオブジェク

<sup>1</sup> Sugimoto, S., 東アジア公文書館における電子記録の保存と利用の確保, no.14, pp.56-67, EASTICA, 2008年3月

<sup>2</sup> PREMIS 編集委員会「保存メタデータのための PREMIS データ辞書」v.2, 2008年3月, <http://www.loc.gov/standards/premis/v2/premis-2-0.pdf> (2008年7月アクセス)

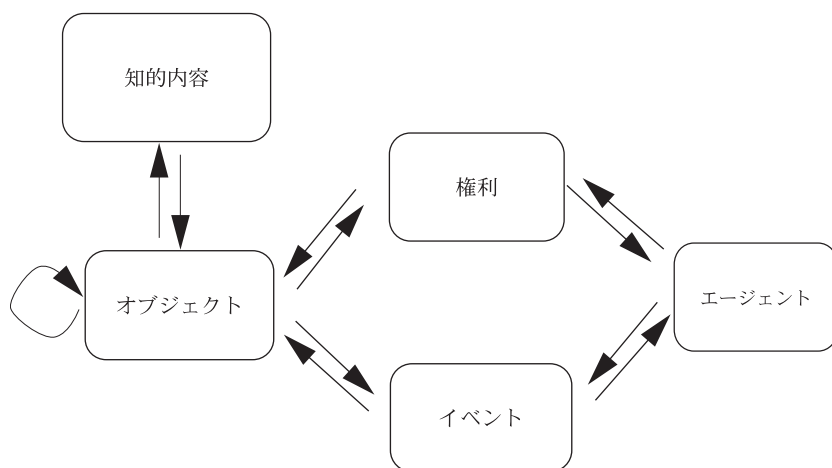


図1 PREMIS データモデル

トに変換することは許される。

### 3. アーカイブとデジタルアーカイビング

国や地方自治体は、文書が紙か電子形態かに関わらず、所定の規則や方針に従い、政府の文書や記録を選別、収集、整理、保存しなければならない。また、保存資源を政府や国民の利用に供する責任もある。

公文書のライフサイクルのはじめの段階では、文書はまず担当部署で作成、利用、保管され、所定の期間の後に、保存目的でアーカイブに転送される。アーカイブは主に文書のライフサイクルの最後の段階を担当する。文書に関する情報（すなわち、文書のメタデータ）は文書のライフサイクルのはじめの段階から作成されるべきである。これはアーカイブが文書を適切に保存整理するために非常に重要な要件である。

行政機関には、情報技術を活用して機能やサービスを強化し、電子政府機能を高度化することが求められている。効率の良い文書情報管理は電子政府機能の強化の基本要素の一つである。情報利用環境を改善することは一般市民にも行政機関の職員にも重要なことである。

現在、ほとんどすべての政府文書はデジタル環境で作られ、ネットワーク情報環境で利用されている。電子政府機能が発達すればするほど、より多くの政府情報資源がより高度なデジタルネット

ワーク環境で作成利用されるようになる。このことは、デジタルアーカイブの実現への強い要望に結びついている。

2007年秋に東京で開催された EASTICA 2007における中国、韓国、モンゴル、日本の国別報告では東アジアにおけるネットワーク情報環境のためのデジタルアーカイブについて色々なモデルや考え方が発表された<sup>3</sup>。

韓国国立公文書館は、公文書の全ライフサイクルを扱う高度な総合公文書管理システムについて報告した。中国国家档案局はアーカイビングに物理的媒体とデジタル媒体の両方を使う公文書の保存方針を報告した。日本の国立公文書館は主にライフサイクルの最終段階のみ担当する機関の視点からデジタル資源をアーカイブするという考え方について報告を行った。

政府部門における文書のライフサイクル全体に渡る記録管理の基本的枠組みは、効率の良いアーカイビングの記録管理には重要な問題である。他方で、政府部門によって記録管理要件が異なることから、理想のシステムを実現するのは容易ではない。例えば、文書形態は、国のアーカイブ方針に基づくのではなく、各政府部門やプロジェクトの方針に応じて選択される。公文書館は様々な方針のもとで作成される様々な文書の保管保存のために、情報技術を存分に活用して、アーカイブを構築していかなければならない。

過去の資料に自由にアクセスできることは多くの人々にとって共通の大きな利点である。たとえば、国立公文書館のアジア歴史資料センターが提供する公文書のデジタルコレクションは、貴重な歴史文書のデジタルアーカイブの役割を理解す

<sup>3</sup> 東アジア公文書館における国別/地域別報告、no.14, pp.68-117, EASTICA, 2008年3月（中国、韓国、モンゴル、日本、マカオからの報告）

るための非常に好例である<sup>4</sup>。同センターは明治時代から第二次世界大戦終了、つまり、1860年代末から1945年までの日本の政府文書を電子化して作成した膨大なデジタルコレクションを提供している。このサービスは、当該期間における、日本の内閣、外務省、陸海軍により記録された文書1,400万ページを超すデジタル画像を提供している。このコレクションは歴史文書のデジタルアーカイブが歴史研究者のみならず家庭や学校の一般利用者にもいかに有用かを示してくれている。

#### 4. デジタルアーカイビングと保存における問題

以下ではデジタルアーカイビングと保存に関わる基本的な問題を幾つか取り上げる。公文書館だけに特有の問題もあれば、図書館や公文書館に共通する基本的問題もある。保存された内容の安全といったよく知られた問題や基本的な保存戦略としてのエミュレーションやマイグレーションは、既に多くの機会に考察されているので、ここでは扱わない。ここに取り上げる問題は主に、国立公文書館におけるデジタルアーカイビングに関する委員会での議論で出たものである。

##### 4.1 文書及び記録の保存用識別 - どんな実体を保存すべきか？

紙の資料のアーカイブの場合、物理的オブジェクト、つまり、紙とマイクロフィルムを収集保存する。デジタルアーカイブの場合、デジタル文書の柔軟性や多様性から保存すべき実体を定義するのは容易ではない。以下、問題例を幾つか挙げる。

- (1) 複雑な構造を持つ文書の問題：デジタル文書にはハイパーリンクが含まれていることが多い。ひとつのハイパーテキスト文書に含まれるハイパーリンクの中には、文書内 (intra-document) のリンクもあれば文書間 (inter-document) のリンクもある。例えば、文書内ハイパーリン

クで相互接続された複数ファイルで構成された文書は単一の資源として扱うべきである。そのため、単一文書の全体を把握するために、文書内リンクと文書間リンクを識別しなければならない。

- (2) 動的文書の作成上の問題：新しいコンテンツ管理環境においては、文書をデータベースに保存されたデータから作成することがよく行われる。ここから、保存すべきはデータベースから作成された文書であるのか、あるいは、その内容自体、つまりデータベースに保存されたままのデータであるのかという素朴な疑問が生じる。
- (3) システムと内容の依存性：デジタル文書は一般にシステム環境と内容に依拠している。例えば、ワープロ文書であればブラウジングに適したワードプロセッサが必要である。文字コードやフォント、XML形式に関する情報を各々保存しなくてはならないので、プレーンテキストやXMLテキストでさえソフトウェアに依拠していないわけではない。

##### 4.2 文書と記録の特徴の識別 - どんな特徴を保存すべきか？

情報技術が常に進展を続けていることから、どんな形態の文書であっても100年間はもちろんたとえ10年間であっても、単純に保存可能であると考えるのは現実的ではない。紙の文書やマイクロフィルムはデジタル文書より安全であると考えられているので、アーカイブの中にはデジタルコンテンツを印刷して印刷資料を保存しているところもある。しかしながら、この場合は、ハイパーリンクやスプレッドシートの計算機能といったデジタルコンテンツの重要な機能が失われる可能性がある。

電子文書の保存のために、文書のみならず文書を利用するシステム全体を保存することが理想ではある。しかしながら、それが可能であると考えすることは現実的ではない。優れたエミュレーションシステムやマイグレーションシステムがあった

<sup>4</sup> アジア歴史資料センター、<http://www.jacar.go.jp/> (2008年7月アクセス)



としても、それが完璧であると考えすることはできない。多くの場合、最初の形態からより保存に適した形態に文書を変換することが必要であろう。この変換により文書の機能面の損失や見た目や使い勝手（Look-and-Feel）に違いが生じる可能性がある。このように、デジタル文書の保存は原本が持つ機能が多少は損なわれることを暗に認めなければならない。アーカイブは機能や情報の損失に関する基準と、そうした基準に従い適当な保存形態を選択する指針を用意しなければならない。

#### 4.3 相互運用性

どんなアーカイブであっても、資源の収集、保持、保存、提供方法のための独自の方針を用意している。一方、アーカイブ同士の相互運用性は、サービスの質を高めるうえで非常に重要な問題である。例えば、アーカイブ横断型の検索、協調的アーカイブ、付加価値サービスのための共通プロトコルやそれに基づくアプリケーションプログラムインターフェース（API）などである。一般に、政府の各部門は、それぞれの分類制度や組織体系を使って、その要件に従い、各部門についての情報資源整備方針を調整する必要がある。その一方、部門間での相互運用性にも留意する必要がある。このことは文書の保存にも影響を及ぼすことになる。

メタデータのない電子文書の利用は困難である。長期的なメタデータの利用性を保証すること、言い換えれば、時を越えたメタデータの相互運用性を保証することは、電子文書の保存上の重要な問題である。そのため、メタデータ要素の定義、属性語彙、属性値記述形式、そして、メタデータの基本作成ルールなど、いわゆるメタデータスキーマを適切に保存しなければならない。

#### 4.4 アーカイブ機能の分割 - 収集、保存、アクセス

オープンアーカイバル情報システム（OAIS）は保存リポジトリの基本的枠組みを定義している。

この枠組みは特定の応用領域やコンテンツとは関係なく定義される。一般に、ビットレベルでの長期保存を前提とすることで、信頼できる保存リポジトリを特定の応用領域とは関係なく構築することができる。しかしながら、資源を収集整理する機能や、保管資源へのアクセスを与える機能は、アプリケーションドメインに規定された要件を満たす必要がある。

図2は、保存機能性を収集及びアクセス機能と分けたアーカイバルシステムの基本モデルである。特に小規模なアーカイブの場合は、保存レポジトリ機能を共有することで全体のコストを下げるのできるのではないかと考えられる。

#### 4.5 ウェブアーカイビング（ファイル保管）

多くの図書館や公文書館が熱心にウェブ資源のアーカイビングに取り組んでいる。そうしたウェブアーカイブは、主に既にウェブ上で公開されている資源を収集蓄積しているものである。ウェブアーカイビング技術は公開された資源のみならず、イントラネット環境など、外部からは見えない組織内部で提供される資源にも適用可能である。その場合、アーカイブは資源収集にあたり、文書提供者と協調的に作業を進めることができるし、また協調すべきである。そうした協調的資源収集は、収集された文書の内容の一貫性の保存を促すであろう。このように、Web技術を利用して組織内で提供される文書をアーカイブすることは重要な問題である。

#### 4.6 アクセシビリティの問題

アーカイブは、文書の主たる作成者ではないが、文書や記録のアクセシビリティ、すなわちすべての利用者が内容にアクセスできることを保証することにも注意を払わなければならない。たとえば、原本の文書がアクセシビリティ指針を満たさず設計となっても、保存のための変換時に指針を満たさなくなってしまうこともある。また、障害者へのアクセシビリティを確保するために利用する

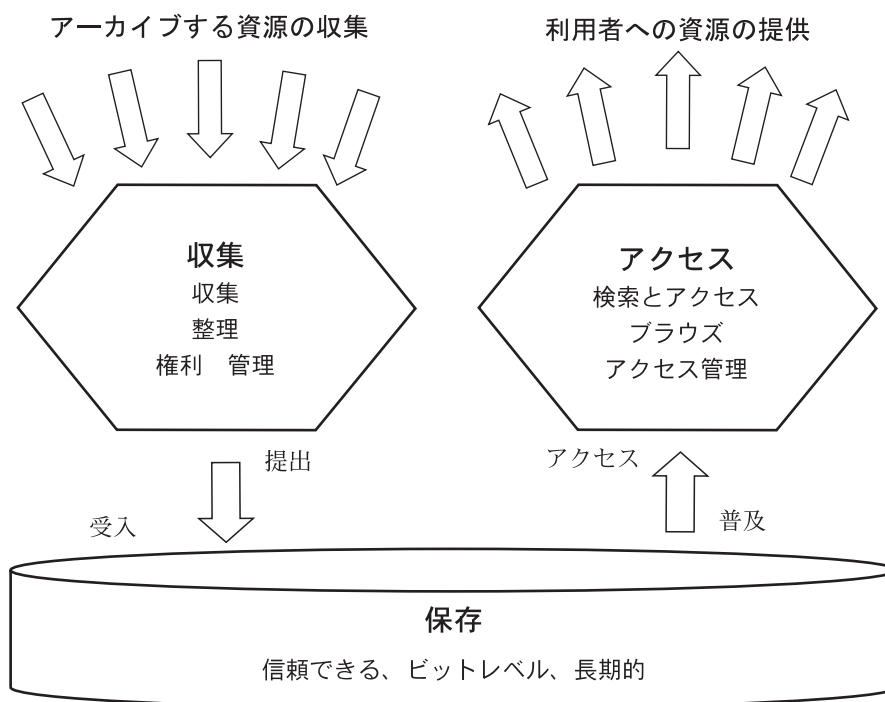


図2 アーカイブシステムの概念的枠組み

ソフトウェアが失われる場合もある。障害の有無に関わらず保存された文書へのアクセスを可能にすることは忘れてはならない課題である。

## 5. まとめ

デジタル資源の保存は、難しい問題として広く知られている。問題は技術分野のみならずアーカイブの管理運営方針や権利管理等にも関連する。我が国においても、公文書の保存に責任を持つ機関はデジタルアーカイブやデジタル資源の長期的利用は重要な問題として広く認識し、将来に向けた努力を重ねている。たとえば、国立国会図書館は電子資料の長期保存とウェブアーカイビングについて調査を行っている<sup>5</sup>。内閣府が主催して始められた公文書の電子管理、移管、保存に関する調査グループにおける討論をもとに<sup>6</sup>、政府のデジタル文書の保存に関する実行可能性調査も進んだ。今後、こうした調査や検討の結果に基づきいろいろな方策が進められていくことになる。

<sup>5</sup> 国立国会図書館、「デジタル情報の長期的な保存と利用」<http://www.ndl.go.jp/aboutus/preservation.html> (日本語、2008年7月アクセス)

デジタル保存やデジタルアーカイブの長期的利用は重要であり、私たちの未来に大きな問題を投げかけている。私たちの貴重な文化、歴史遺産を将来に向け保存するために、アーカイブや記録保管機関同士で技術、技能、経験を共有する協力体制が求められていると強く感じる。公文書の電子的保存は国のみならず自治体やさまざまな企業体にとっても重要な話題である。より規模の小さなコミュニティが保管している貴重な資源を保存するためには、分担や協調によるコスト削減、効率的なアーカイブ構築が重要である。

電子文書の保存は決して派手な仕事ではない。しかしながら、デジタル情報技術に支えられた現代の社会が生み出す様々な情報資源を将来の社会に伝える重要な仕事である。

末筆ながら、国立公文書館をはじめとして、電

<sup>6</sup> 公文書等の適切な管理、保存、利用に関する懇談会(内閣府)、「中間段階における集中管理及び電子媒体による管理、移管、保存に関する報告書」2006, p32. <http://www8.cao.go.jp/chosei/koubun/kondankai14/houkoku.pdf> (日本語、2008年7月アクセス)

子公文書、デジタル文書の保存に関して考える機会を与えてくださった方々に感謝の意を表したい。

発表者略歴

京都大学・工学部・情報工学科、同大学院情報工学専攻修了（工学博士）、京都大学工学部、図書館情報大学図書館情報学部を経て、現在、筑波大学図書館情報メディア研究科教授、筑波大学知的コミュニティ基盤研究センター・センター長。

国立国会図書館納本制度審議会専門委員（ネットワーク系電子出版物の収集に関する制度のあり方について、2002年～2005年）、Dublin Core Metadata Initiative の評議委員会（Board of Trustees）、諮問委員会委員（Advisory Board）、Joint Conference on Digital Libraries (JCDL) および International Conference on Asian Digital Libraries (ICADL) の Steering Committee 委員、JCDL 2007, ICADL 2007, DC 2001のプログラム共同委員長





## Session 3

# 危機とアーカイブズ

## Crisis and Archive

### ◆プログラム◆

モデレーター：高山 正也 国立公文書館理事

14:30 15:00 『山一証券資料』について On the Materials of Yamaichi Securities

伊藤 正直 東京大学経済学部教授

15:00 15:30 『日本石炭産業関連資料コレクション』 - 目録データベースの公開と今後の課題 -

Japanese Coal Industry Collection at Keio University Library: catalogue data-  
base and research perspective

杉山 伸也 慶応義塾大学経済学部教授

15:30 16:00 原子力産業における重要記録の管理

Managing Critical Records in the Nuclear Power Industry

山下 貞麿 日本レコードマネジメント株式会社コンサルタント代表

16:00 16:30 記録なくして、歴史なし - 沖縄における記録再構築の試み -

No Records, No History? Then what would happen to Okinawans?

仲本 和彦 財団法人沖縄県文化振興会公文書管理部

16:30 17:00 質疑応答・意見交換 Discussion and Q&A

